

No. of Printed Pages : 8

A10.1-R5 : DATA SCIENCE USING PYTHON

DURATION : 03 Hours

MAXIMUM MARKS : 100

OMR Sheet No. :

--	--	--	--	--

Roll No. :

--	--	--	--	--	--

Answer Sheet No. :

--	--	--	--	--	--

Name of Candidate : _____ ; Signature of Candidate : _____

INSTRUCTIONS FOR CANDIDATES :

- Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.
- Question Paper is in English language. Candidate has to answer in English language only.
- There are **TWO PARTS** in this Module/Paper. **PART ONE** contains **FOUR** questions and **PART TWO** contains **FIVE** questions.
- **PART ONE** is Objective type and carries **40** Marks. **PART TWO** is Subjective type and carries **60** Marks.
- **PART ONE** is to be answered in the **OMR ANSWER SHEET** only, supplied with the question paper, as per the instructions contained therein. **PART ONE** is **NOT** to be answered in the answer book for **PART TWO**.
- Maximum time allotted for **PART ONE** is **ONE HOUR**. Answer book for **PART TWO** will be supplied at the table when the Answer Sheet for **PART ONE** is returned. However, Candidates who complete **PART ONE** earlier than one hour, can collect the answer book for **PART TWO** immediately after handing over the Answer Sheet for **PART ONE** to the Invigilator.
- **Candidate cannot leave the examination hall/room without signing on the attendance sheet and handing over his/her Answer Sheet to the invigilator. Failing in doing so, will amount to disqualification of Candidate in this Module/Paper.**
- After receiving the instruction to open the booklet and before answering the questions, the candidate should ensure that the Question Booklet is complete in all respects.

DO NOT OPEN THE QUESTION BOOKLET UNTIL YOU ARE TOLD TO DO SO.

PART - ONE

(Answer all questions; each question carries ONE mark)

1. Each question below gives a multiple choice of answers. Choose the most appropriate one and enter in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

1.1 Which type of Programming does Python support ?

- (A) Object-Oriented Programming
- (B) Structured Programming
- (C) Functional Programming
- (D) All of the above

1.2 All keywords in Python are in _____.

- (A) Capitalized
- (B) lower case
- (C) UPPER CASE
- (D) None of the above

1.3 Which of the following functions is a built-in function in python ?

- (A) factorial()
- (B) print()
- (C) seed()
- (D) sqrt()

1.4 What is the purpose of NumPy in Python ?

- (A) To do numerical calculations
- (B) To do scientific computing
- (C) Both (A) and (B)
- (D) None of the above

1.5 Amongst which Python library is similar to Pandas ?

- (A) SciPy
- (B) SymPy
- (C) NumPy
- (D) None of the above

1.6 What will be the minimum number of arguments require to pass in Pandas series ?

- (A) 2
- (B) 3
- (C) 4
- (D) None of the above

1.7 Amongst which of the following is/are used to analyze the data in pandas ?

- (A) Dataframe
- (B) Series
- (C) Both (A) and (B)
- (D) None of the above

1.8 Which type of machine learning algorithm falls under the category of "unsupervised learning" ?

- (A) Linear Regression
- (B) K-Means Clustering
- (C) Decision Trees
- (D) Random Forest

1.9 An artificially intelligent car decreases its speed based on its distance from the car in front of it. Which algorithm is used ?

- (A) Naïve-Bayes
- (B) Decision Tree
- (C) Linear Regression
- (D) Logistic Regression

1.10 Which function is used to draw points (markers) in a diagram ?

- (A) write()
- (B) draw()
- (C) plot()
- (D) paint()

2. Each statement below is either TRUE or FALSE. Choose the most appropriate one and enter your choice in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

2.1 Binomial Distribution is a Discrete Distribution.

2.2 A series object is size mutable.

2.3 Matplotlib is not open-source and free.

2.4 The ix_ function can be used to combine different vectors.

2.5 A Dataframe object is value mutable.

2.6 Indexing in Series is not similar to that for NumPy arrays.

2.7 Tkinter is the inbuilt python module that is used to create GUI applications.

2.8 Maximum likelihood estimation gives us not only a point estimate, but a distribution over the parameters that we are estimating.

2.9 The Box plot are unable to show outliers.

2.10 DataFrame accepts many different kinds of input.

3. Match words and phrases in column X with the closest related meaning/word(s) /phrase(s) in column Y. Enter your selection in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

	X		Y
3.1	A Python library for data manipulation and analysis	A	Coefficient of variation
3.2	The function is used to create an identity matrix in NumPy.	B	Pandas
3.3	A Python interactive visualization library for large datasets that natively uses the latest web technologies.	C	Class
3.4	The ratio of the mean to the standard deviation	D	Supervised Learning
3.5	Random variable	E	eye()
3.6	Statistical computing and analysis	F	Stochastic
3.7	Scientific routines that work on top of NumPy	G	Data Science
3.8	User-defined data type	H	SciPy
3.9	A multidisciplinary which involves extraction of knowledge from large volumes of data that are structured or unstructured	I	R Language
3.10	An algorithm uses input-output pairs to learn patterns and make predictions.	J	Bokeh
		K	Drawing
		L	Guido van Rossum
		M	Data

4. Each statement below has a blank space to fit one of the word(s) or phrase(s) in the list below. Enter your choice in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

A	lambda	B	Def	C	.py
D	Plotting	E	String	F	Accuracy
G	Guido van Rossum	H	Pip	I	Serializing
J	itercolumns()	K	Func	L	James Gosling
M	Set				

- 4.1 _____ developed Python Programming Language.
- 4.2 _____ is the extension of the Python file.
- 4.3 Python supports the creation of anonymous functions at runtime, using a construct called _____.
- 4.4 Matplotlib is a _____ library for the Python programming language.
- 4.5 The _____ is not an iterative function for dataframe.
- 4.6 Pickling is the process of _____ a Python object, that is, conversion of a Python object hierarchy into a byte stream.
- 4.7 _____ is a standard package management system used to install and manage Python software packages.
- 4.8 _____ data type in Python is used to store a collection of items, where each item is unique and unordered.
- 4.9 _____ data type in Python represents a sequence of characters.
- 4.10 _____ is commonly used for evaluating classification models.

PART - TWO

(Answer any FOUR questions)

5. (a) What is NumPy, and what are its primary features ? Provide an example of how NumPy is used in data manipulation.
- (b) Define a structured array with two columns. The first column contains the product ID, which can be defined as an int32. The second column shall contain the price for the product. How can you print out the column with the product IDs, the first row and the price for the third article of this structured array ? Figure out a data type definition for time records with entries for hours, minutes and seconds. (6+9)
6. (a) Explain the purpose of pandas in data analysis. How does it differ from NumPy and what are some key DataFrame operations ?
- (b) What is the importance of framing the problem correctly in the data science process ? Provide examples of how framing the problem impacts the outcome of a data science project. (8+7)
7. (a) What is matplotlib, and how is it used in data visualization ? Provide an example of a plot created using matplotlib.
- (b) What is fancy indexing in NumPy, and how does it differ from basic indexing ? (9+6)

8. (a) Discuss about the operations of Sequences in Python with proper example.
- (b) Write a Python program that creates a list of numbers and then calculate the sum of all the numbers in the list. (10+5)
9. (a) Explain the purpose of the following Tkinter widgets :
- (i) Canvas
 - (ii) Label
 - (iii) Entry
 - (iv) Frame
 - (v) Checkbutton
- (b) Discuss the significance of mean, median, mode and standard deviation in summarizing and describing data. (10+5)

- o o o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK